The Community Software Repository from XSEDE

JP Navarro
Argonne National Laboratory
MCS Division, TCS Building 240
Lemon, IL 60439
(630) 252-1233
navarro@mcs.anl.gov

Richard Knepper
Indiana University
2709 E 10th Street
Bloomington, IN 47408
(812) 855-9574
rknepper@iu.edu

Lee Liming
University of Chicago
5801 South Ellis Avenue
Chicago, Illinois 60637
(773) 702-1234
lliming@uchicago.edu

Maytal Dahan
Texas Advanced Computing Center
University of Texas
Austin, TX 78758
(310) 429-9419
maytal@tacc.utexas.edu

David Lifka
Cornell Center for Adv. Computing
Frank H. T. Rhodes Hall, Hoy Road
Ithaca, NY 14853
lifka@cac.cornell.edu

Craig A. Stewart
Indiana University
2709 E. Tenth Street
Bloomington, IN 47408
stewart@iu.edu

ABSTRACT

The Extreme Science and Engineering Discovery Environment (XSEDE) [1] aims to connect cyberinfrastructure (CI) resources, software, and services. By bringing together advanced digital infrastructure, expert support, and training services, XSEDE enables scholars, researchers, and engineers to participate in multidisciplinary collaborations that seamlessly access advanced computing resources and share data to tackle society's grand challenges.

Software enables this endeavor by connecting infrastructure elements into an integrated CI and by enabling users to access that CI. Since XSEDE is not funded to develop new integrating and access software it must engage with a community of developers, integrators, vendors, users, operators, and funding agencies to document requirements; develop, share, and deploy software; and provide software documentation, training, and support. To manage these software related activities XSEDE has adopted tools. Since these are community activities and not just XSEDE activities, XSEDE is now sharing its tools thru a new **Community Software Repository (CSR)** service. By providing an online service where software providers and consumers can engage in these software activities we will accelerate collaboration and expand software based CI capabilities in XSEDE. This paper describes the CSR vision and strategy, current capabilities, future plans, and related XSEDE efforts.

Categories and Subject Descriptors

K.6.3 [**Software Management**]: Software requirements and use cases (D.2.1), engineering best practices (D.2), software quality assurance (D.2.9), testing and debugging (D.2.5), software distribution and support (D.2.7).

General Terms

Management, Documentation, Design, Security, Human Factors, Standardization, Verification.

Keywords

Software requirements, engineering activities, priorities, technical reviews, quality assurance, distribution, publishing, discovery, support, and collaboration.

1. Introduction

The mission of the XSEDE Community Infrastructure (XCI) team is to facilitate interaction, sharing and compatibility of all relevant software and related services across the national CI community building on and improving on the foundational efforts of XSEDE.

Our vision is a vibrant and diverse software developer, integrator, operator, funding agency, and user community collaborating to accelerate resource, software, and software based service integration into the national research CI. The CSR was born out of the need to promote open collaboration, information sharing, software sharing, and discussion between software providers and software consumers.

2. CSR STRATEGY

XSEDE, federated service providers (SPs) [2], and software partners integrate software tools that enable users to access and use an integrated CI. XSEDE uses best practices involving software providers and consumers to ensure that software addresses the most pressing user needs. These best practices are enabled by the CSR and include: engaging users in documenting requirements and use cases, documenting capability gaps, involving users and other stakeholders in setting capability gap priorities, involving stakeholders in reviewing software designs and security, quality assurance testing, sharing software, enabling software discovery, coordinating deployment and user support, and providing software documentation and training.

These activities involve both software provider and software consumer stakeholders. The CSR provides these stakeholders easy access to these best practices by enabling sharing, discovering, and discussing of 1) user requirements, use cases, and priorities; 2) CI integration activities; and 3) software available to users, developers, and operators. The CSR enables software sharing and discovery at all life-cycle stages: whether it is operational and ready to use, whether it first needs to be built or installed, or if it is under development. The CSR should also

enable software sharing in all its formats: in a conventional package, in a VM, a container, as a build recipe, as operational SaaS, on a file-system on an HPC machine, or in a source repository.

To help the community adopt the CSR web portal and to evolve it to address the community's challenges we will engage face-to-face with **software developers**, **integrators**, **operators**, and **consumers** thru relevant events and meetings, and provide documentation and training on using the CSR. We will also look at both technical and human adoption barriers [3].

2.1 Part of a Broader Community

The research software community has a very diverse set of software sharing, discovery, and information access requirements. Constrained by XSEDE's mission, the CSR can't provide all the user facing interfaces that software consumes in the broader community may need. Similarly, we can't integrate all software related best practices into the CSR toolkit. We are therefore initially focusing on the basic functionality to support XSEDE's own needs, and the needs of our direct software partners, while still providing some basic functionality to others in the community. To enable the community to build custom interfaces that may not be useful to XSEDE and to enable XSEDE to build it's own custom interfaces we will develop a web API with access to all CSR information. Initial elements of this API are already in use by the XSEDE User Portal (XUP) to build the software discovery interfaces for XSEDE users.

We further recognize that there are many software management repositories and tools in use in the community that satisfy project specific, institutional, research collaboration, and other needs. The CSR is not a replacement for these community specific repositories and tools. The CSR offers a public location to share research CI software information in a consistent format that enables discovery across the broadest possible community. Is a sense a **Research Community Software Yellow Pages**. To enable this "web of software information" the CSR provides the ability to reference external software repositories and catalogs where more specialized or community specific software information is available. This will enable users who discover software thru the CSR to navigate to the repositories that CSR content came from. All the software that XSEDE integrates for federated SPs, campuses, and users is discoverable thru the CSR.

The next three section describe how the CSR supports three key customer communities: developers, operators/integrators, and end users.

2.2 Enabling Software Developers

For the **software developers** community the CSR enhances:

- Sharing requirements and use cases that are enabled by their software
- Publishing use case support gaps
- Gathering input on which gaps have the highest priority
- Publishing the availability of packaged software in any form
- Publishing the availability of operational ready-to-use software in any form
- Engaging in open discussion on software requirements, priorities, deployment, and use

Software developers can pick a-la-carte which of these CSR capabilities they want to leverage. Software developers can consult with the XCI team on how to effectively use the CSR capabilities they choose to use or how those capabilities could be enhanced to better support their software sharing needs.

2.3 Enabling Operators and Integrators

For the **software operators/integrators** community the CSR enhances:

- Discovering software that enables use cases that they want to support
- Discovering resource integration, federation, and interoperability options enabled by software [4]
- Discovering the availability of packaged software in any form that they can adopt and deploy
- Publishing the availability of operational ready-to-use software in any form (which they may have discovered in the CSR from a developer)
- Engaging in open discussion about requirements, priorities, deployment, or use of software

Software operators and integrators can pick a-la-carte which of these CSR capabilities they want to leverage. They can consult with the XCI team on how to effectively use the CSR capabilities they choose to use or how those capabilities could be enhanced to better support their software sharing needs.

2.4 Enabling End Users

As part of the CSR effort we are engaging with **software consumers** to ensure that the CSR provides them the software discovery capabilities they need. We will enhance the CSR to address discovery requirements, and work with providers so that they can publish the information needed to enable that discovery.

From the CSR perspective, users are anyone who is looking for software that they don't have. As discussed in the previous section, this includes operators and integrators who obtain software from developers and other integrators that they will install and support for their end-users.

The CSR offers software consumers a tool to:

- discover software that enables use cases of interest,
- discover ready-to-use software and documentation,
- discuss requirements and use cases, priorities, and experiences with software, and
- navigate to specialized software repositories.

3. CSR INFORMATION TYPES

The CSR contains the following software related information.

Use Cases: How can software providers describe what software does? How can software consumers discover new software based on what they want to do? Use cases are the approach that XSEDE successfully adopted to help answer both questions, and to build an implementation agnostic bridge between software providers and consumers. Use cases describe what users can do. They are the requirements or expectations of a software consumer that a software provider must satisfy. The CSR includes a Use Case Registry that contains use case descriptions, target audience summaries, community sizes, use case review status, use case area/sub-area, last proposed and last approved dates, and related document pointers.

Capability Delivery Plans and Priorities: Use Case support

varies across different infrastructures. Capability Delivery Plans (CDPs) are concise documents that describe what aspects of a Use Case are supported by a particular infrastructure, which aspects are not supported, and an infrastructure project's plans to fill those gaps. Infrastructures have to periodically prioritize which use case support gaps they are going to work on. The CSR includes the capability to survey a community of users for their input on which use cases should be prioritized to drive activities that fill support gaps.

Software Activities: Software projects propose, prioritize, prepare, and deliver new software products or enhancements to existing products. The CSR includes information about XSEDE software integration activities and how they map to use cases, packaged software, and operational software.

As described above, software comes in many formats and life-cycle availability stages. The CSR contains three types of software information: global, packaged, and operational:

- **Global Software Information** is information that doesn't change based on packaging format and operational status, such as: name, textual description, global categories/tags, and vendor.
- **Packaged Software Information** is information about software that requires action to make it operational. Packages may be in VM, container, RPM, tar, build recipe, or other format that can be instantiated. Package information includes repository pointers; installation / build / provisioning instructions; format; and package support contact information.
- **Operational Software Descriptions** is information about "ready to use" operational software available by command line (on an SP resource[5]) or thru a network interface (SaaS, portal, gateway, etc.). Information includes: how to access and use the software, and operational support contact information.

Resources: Multiple software components can coexist on a single infrastructure resource, such as an HPC/HTC/Viz resource, integrated into a science gateway, or in some other network accessible service. The CSR is linked to XSEDE's Resource Description Repository (RDR) which describes the CI resources where operational software is available.

Discussions: In the CSR we will include open community discussions on software driving use cases, priorities, delivery plans, delivery activities, designs, testing, deployment, and use.

4. CURRENT CSR CAPABILITIES

XSEDE introduced the CSR in October of 2016 and expanded it in January of 2017 and April of 2017.

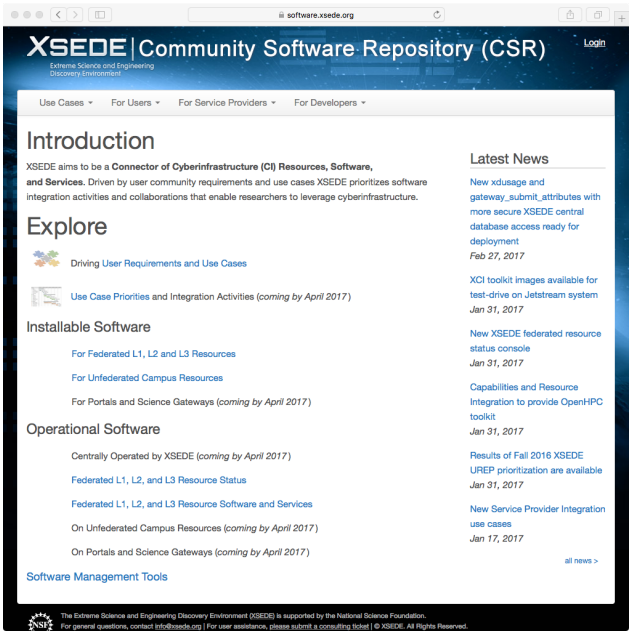

Figure 1. The CSR main page

4.1 Initial Fall 2016 Features

The XSEDE system is explicitly intended by the NSF to be open and extensible. It must not only respond to the high-end computing needs of the research community, it must do so in a way that is transparent to others and that allows others to offer their own solutions to unmet community needs. The CSR contributes to this goal by offering a catalog of community needs that we have identified while working with the scientific disciplines and resource operators that XSEDE supports. Each user need is referred to as a *use case*, and the CSR provides a *use case repository*. The use case repository contains a set of ~70 use cases that were identified from 2012 through 2016, when XSEDE formulated its foundational architecture. [6] The number of use cases is continuously growing as XSEDE staff identity new usage scenarios and user needs.

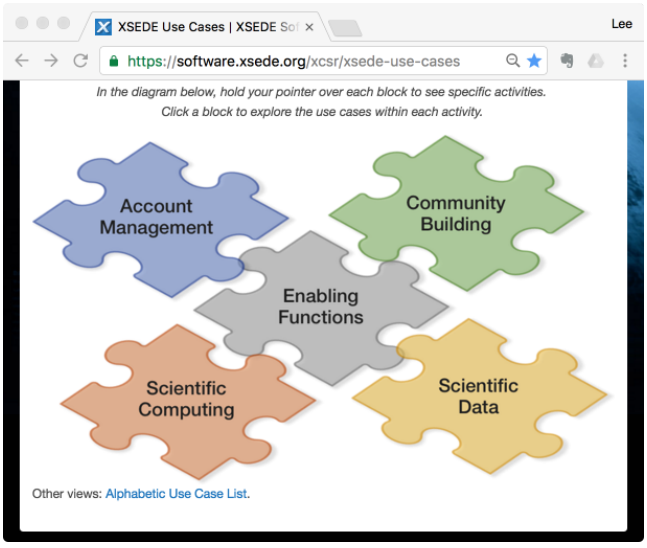

Figure 2. The CSR use case repository

Each entry in the repository includes a description of something someone needs to do. (This description is the use case.) Many entries also include a pointer to an XSEDE capability delivery plan, which details how the current XSEDE system satisfies the use case, which parts may not yet be satisfied, and any plans XSEDE has for closing the gap between current features and the full user need. This information is available to all community members and can be used to identify new features and services that might be of value to the community.

XSEDE gives community members access to a wide range of software. This software includes compilers, applications libraries, development and debugging tools, and other tools that enable users to make effective use of CI. With these tools, users can authenticate and manage security credentials, login interactively to resources, move data between distributed resources within and outside of XSEDE, execute jobs or provision VMs, account for allocations usage, etc.

Information about the software enabling these capabilities is provided in the CSR. Using this information, SPs, campus CI providers, and other community members can learn about these software tools and deploy them on their own systems. The CSR includes all the software tools that XSEDE has integrated and made available in packaged formats for community members. The XSEDE CRI team can also provide consulting on implementation details for local CI implementations.

4.2 Winter 2017 Features

Continuing the theme of an open and extensible system, the CSR provides a record of the decision-making processes that have resulted in new system features. Twice each year, XSEDE's User Requirements Evaluation and Prioritization (UREP) working group reviews the new use cases and capability delivery plans that have been identified by the XCI team. These reviews are conducted online and the results are stored for later reference.

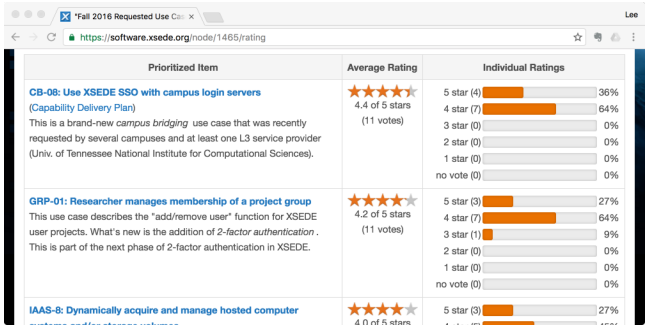

Figure 3. Results of use case prioritization in the CSR

XSEDE uses the results to guide its own work on system features, but the data is available in the CSR for members of the user community and potential SPs to use for their own purposes.

To enhance the ability of XSEDE SPs to follow the steps required integrate into the XSEDE environment, and to enable SPs and the XSEDE integration coordinator to track integration status, the CSR includes a **Resource Integration Console** [7] showing completed and outstanding steps for integrating CI HPC, HTC, visualization, and storage resources into the XSEDE environment. This console includes details about required and optional software that SPs can use and/or install to complete the integration.

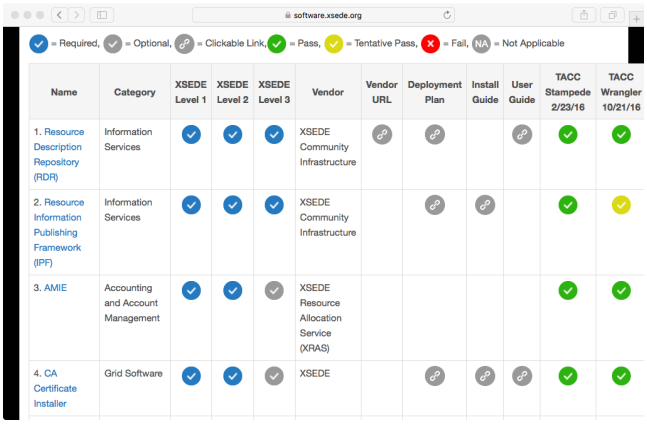

Figure 4. The CSR Integration Console

To enhance the ability of XSEDE SPs to track operational status, the CSR includes an **Operational Status Console** showing the declared operational status of a resource (from the Resource Description Repository "RDR"), announced outage information, monitoring results, and resource publishing status [8] (includes whether a resource has up-to-date software availability information).

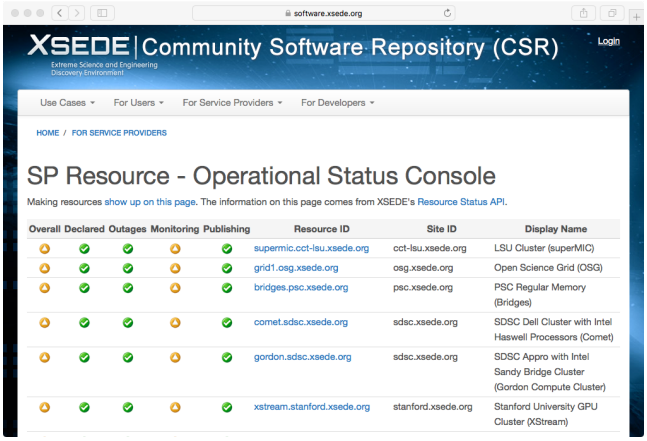

Figure 5. The CSR Operational Console

4.3 Spring 2017 Features

During the Spring of 2017 the CSR is being enhanced to include the following expanded software information.

a) A **Global Software Descriptions** repository that includes information about software that is independent of its status or availability, including: software description, vendor details, and global tags and categories. This feature will include web forms for entering Global Software Descriptions and may introduce a vetting process so that CSR administrators can review and correct Global Software Description quality issues.

b) Methods for XSEDE and Science Gateways to manually enter **Operational Software Descriptions** using web forms. This will enable science gateways and XSEDE itself to advertise user and developer facing software and service information.

5. CSR Future

XSEDE has planned a series of enhancements for RY7 (May 2017 thru April 2018) that will support ongoing efforts to improve the software sharing process. Beyond what it can do within its mandate and within baseline funding XSEDE is pursuing Project Improvement Funds and separate NSF software funds to continue to expand the usefulness of the CSR to the broader research software community. Areas of future attention include:

- The campus CI toolkits which are available in the CSR will be extended to include an OpenHPC-based toolkit as well as a data science toolkit, as that has been highly requested by the community.
- The CSR will also be extended with links to a larger range of virtual machine and containerized applications that can be implemented easily within existing infrastructure.
- Enabling community software providers to automatically advertise the software they have placed in XSEDE Community Software Areas (CSAs).
- Comprehensive Use Case views that enable use case discovery and navigation to all related information.
- Comprehensive Software views that enable software discovery and navigation to all related information, such as packaged and operational availability, and use cases.
- Engaging the community to add their software information to the CSR.

5.1 Plans contingent on additional funding

Currently under proposal via the XSEDE Project Improvement Fund system is a plan to provide initialization scripts and instantiation instructions for popular virtual machines offered on the Jetstream research cloud. The team will distill a set of VM's offered on the Jetstream system into sets of instructions (Openstack and Heat configurations, with cloud-init for post-instantiation configuration), so that they can be implemented in other cloud infrastructures. These recipes will be tested on the Bridges system at PSC before release to the CSR. A follow-up proposal to NSF is planned to cover additional development of VM recipes for further implementation.

5.2 XSEDE User Portal Plans

The initial goal of XUP was to provide XSEDE community and users comprehensive software information. Using the CSR API the XUP team developed an easy-to-use user interface to search for XSEDE software across sites. The initial interface enables the ability to gather automated software information from CSR in addition to add more static metadata via an administrative interface. Users have the ability to search the software catalog and filter by science domain, service provider, or specific system.

The XUP team will continue to work alongside CSR to continue to promote and offer software and service information to the end-user.

6. ACKNOWLEDGMENTS

This document was developed with support from National Science Foundation (NSF) grant OCI-1053575. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357.

Figure 1. Insert caption to place caption below figure.

7. REFERENCES

[1] XSEDE: https://www.xsede.org/

[2] XSEDE Federation: https://www.xsede.org/xsede-federation

[3] Katherine A. Lawrence and Nancy Wilkins-Diehr. 2012. Roadmaps, not blueprints: paving the way to science gateway success. XSEDE '12 DOI: http://dx.doi.org/10.1145/2335755.2335837

[4] Craig A. Stewart, Richard Knepper, James Ferguson, Felix Bachmann, Ian Foster, Andrew Grimshaw, Victor Hazlewood, and David Lifka. What is campus bridging and what is XSEDE doing about it? XSEDE '12. DOI: http://dx.doi.org/10.1145/2335755.2335844

[5] Matthew Hanlon, Warren Smith, and Stephen Mock. Providing resource information to users of a national computing center. XSEDE '13. DOI: http://dx.doi.org/10.1145/2484762.2484826

[6] Bachmann, F., Foster, I., Grimshaw, A., Lifka, D., Liming, L., Reidel, M., Tuecke. S. XSEDE Architecture Overview, version 2.0. Sep. 2014. (http://hdl.handle.net/2142/50274)

[7] XSEDE Service Provider Checklist (v2.0) (http://hdl.handle.net/2142/91024)

[8] Warren Smith, Sudhakar Pamidighantam, and John-Paul Navarro. Publishing and consuming GLUE v2.0 resource information in XSEDE. XSEDE '15. DOI: http://dx.doi.org/10.1145/2792745.2792770

[9] XSEDE Cloud VM Repository (http://hdl.handle.net/2142/90419)